STAT 452 Project 2

**Section 1: ROC curves**

* Wish to display *specificity* and *sensitivity* of a classification model (for example, LDA)
* *Specificity*: percentage of non-defaulters that are correctly identified (true negative rate)
* *Sensitivity*: percentage of true defaulters that are identified (true positive rate)
* ROC curve plots 1-specificity (false positive rate) on the x-axis, and sensitivity on the y-axis. Ideally, we want a high sensitivity and a low 1-specificity
* ROC curve displays the 2 errors over all possible thresholds
* Default classifier (threshold) is 0.5, but can be higher/lower depending on the trade-off of the 2 errors
  + A lower threshold increases both 1-specificity and sensitivity, while a higher threshold decreases both
* Summary table:

|  |  |  |
| --- | --- | --- |
| **specificity** | True Negative Rate |  |
| **1-specificity** (x-axis) | False Positive Rate |  |
| **Sensitivity** (y-axis) | True Positive Rate |  |

* The overall performance of a classifier, summarized over all possible thresholds, is given by the area under the (ROC) curve (**AUC**)
* The larger the AUC, the better the classifier (an ideal ROC curve will hug the top left corner)
* AUC should not be lower than 0.5, because at 0.5 it is the same as a classifier purely predicting by chance

**Section 2: Support Vector Machines (SVM)**